11/PRTS

09/869401
JC18 Rec'd PCT/PTO 27 JUN 2001

Methods and Devices for Coding or Decoding an Audio Signal or Bit Stream

Field of the Invention

The present invention relates to methods and devices for coding or decoding an audio signal or bit stream which are able to perform error-tolerant entropy coding or decoding and in particular error-tolerant Huffman coding or decoding.

Background of the Invention and Prior Art

Modern audio coding or decoding methods, which operate according to the standard MPEG layer 3 for example, are capable of compressing the data rate of audio signals by a factor of 12 for example without causing any noticeable deterioration in the quality of these signals. To obtain such a high data rate reduction an audio signal is sampled, resulting in a sequence of discrete-time samples. As is known in this branch of technology, this sequence of discrete-time samples is windowed using suitable window functions to obtain windowed blocks of temporal samples. A block of temporal windowed samples is then transformed into the frequency domain by means of a filter bank, a modified discrete cosine transform (MDCT) or some other suitable method to obtain spectral values which together represent the audio signal, i.e. the temporal section which consists of the block of discrete-time samples, in the frequency domain. Normally temporal blocks which overlap by 50% are generated and are transformed into the frequency domain by means of an MDCT. Because of the special properties of the MDCT, 1024 discrete-time samples for example always result in 1024 spectral values.

It is known that the receptivity of the human ear depends on the momentary spectrum of the audio signal itself. This dependence is reflected in the so-called psychoacoustic model. Using this model it has long been possible to calculate masking thresholds in dependence on the momentary spectrum. Masking means that a particular tone or spectral portion is rendered inaudible when e.g. a neighbouring spectral region has a relatively high energy. This phenomenon of masking is exploited so as to quantize the post-transform spectral values as coarsely as possible. The aim, therefore, is to avoid audible disturbances in the decoded audio signal while using as few bits as possible to code, or here to quantize, the audio signal. The disturbances introduced by quantization, i.e. the quantization noise, should lie below the masking threshold and thus be inaudible. In accordance with known methods the spectral values are therefore subdivided into so-called scale factor bands, which should reflect the frequency groups of the human ear. Spectral values in a scale factor group are multiplied by a scale factor so as to scale spectral values of a scale factor band as a whole. The scale factor bands scaled with the scale factor are then quantized, producing quantized spectral values. It is of course obvious that a grouping into scale factor bands is not essential. This procedure is, however, used in the standard MPEG layer 3 and in the standard MPEG-2 AAC (AAC = Advanced Audio Coding).

A very important aspect of data reduction is the entropy coding of the quantized spectral values resulting from quantization. A Huffman coding is normally used for this. A Huffman coding entails variable-length coding, i.e. the length of the code word for a value to be coded depends on the probability of this value occurring. As is logical the most probable symbol is assigned the shortest code, i.e. the shortest code word, so that very good redundancy reduction can be achieved

with Huffman coding. An example of a universally known variable-length coding is the Morse alphabet.

In audio coding Huffman codes are used to code the quantized spectral values. A modern audio coder which operates e.g. according to the standard MPEG-2 AAC uses different Huffman code tables, which are assigned to the spectrum according to particular criteria on a sectional basis, to code the quantized spectral values. Here 2 or 4 spectral values are always coded together in one code word.

One way in which the method according to MPEG-2 AAC differs from the method MPEG layer 3 is that different scale factor bands, i.e. different spectral values, are grouped into an arbitrarily large number of spectral sections. In AAC a spectral section contains at least four spectral values, preferably more than four spectral values. The whole frequency range of the spectral values is thus divided up into adjacent sections, where one section represents a frequency band, so that all the sections together cover the whole frequency range which is spanned by the post-transform spectral values.

To achieve a maximum redundancy reduction, a so-called Huffman table, one of a number of such tables, is assigned to each section as in the MPEG layer 3 method. In the bit stream of the AAC method, which normally has 1024 spectral values, the Huffman code words for the spectral values are now in an ascending frequency sequence. The information on the table used in each frequency section is transmitted in the side information. This situation is shown in Fig. 2.

In the case chosen to serve as an example in Fig. 2 the bit stream comprises 10 Huffman code words. If one code word is always formed from one spectral value, 10 spectral values can then be coded here. Usually, however, 2 or 4 spectral values